A CHI-SQUARE STATISTIC FOR GOODNESS-OF-FIT TESTS

by

BU-388-M September, 1971

K. C. Rao and D. S. Robson
Cornell University

Abstract

Chernoff and Lehmann (Ann. Math. Stat., 1954) have shown that the chi-square goodness-of-fit statistic is asymptotically distributed as $\chi^2_{r-s-1} + \lambda_1 Z^2_{r-s} + \cdots + \lambda_s Z^2_{r-1}$ when the r classes are predetermined and the s unknown parameters are estimated by maximum likelihood from the ungrouped data. The Z_i are NIID(0,1) independent of χ^2_{r-s-1} and the λ_j, $0 < \lambda_j < 1$, depend on the s unknown parameters in $F(x;\theta)$. Subsequent papers have shown that this same result applies in the more realistic and useful case where only the number of classes r and their probability content with respect to $F(x;\hat{\theta})$ are predetermined. In either case the joint conditional distribution of the class frequencies $\nu = (\nu_1, \cdots, \nu_{r-1})$, conditioned on $\hat{\theta}$, is asymptotically nonsingular multinormal and the quadratic form $Q_{r-1}(\nu;\hat{\theta},\theta)$ of this conditional distribution is therefore asymptotically distributed as χ^2_{r-1}.

If $F(x;\theta)$ belongs to the Koopman-Pitman family then this quadratic form does not depend on θ; in other cases the substitution of $\hat{\theta}$ for θ still gives $Q_{r-1}(\nu;\hat{\theta},\hat{\theta}) \stackrel{\mathcal{L}}{\to} \chi^2_{r-1}$. Structurally, this statistic takes the form $Q_{r-1}(\nu;\hat{\theta},\hat{\theta}) = X^2 + Q^*_s(\nu;\hat{\theta},\hat{\theta})$ where Q^*_s is a function of the estimated information matrix and

$$Q^*_s(\nu;\hat{\theta},\hat{\theta}) \stackrel{\mathcal{L}}{\to} (1-\lambda_1)Z^2_{r-s} + \cdots + (1-\lambda_s)Z^2_{r-1}.$$

Corresponding results are obtained when maximum likelihood estimators are replaced by other asymptotically normal consistent estimators.

Paper No. BU-388-M in the Biometrics Unit Mimeograph Series, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14850.

A CHI-SQUARE STATISTIC FOR GOODNESS-OF-FIT TESTS

by

BU-388-M September, 1971

K. C. Rao and D. S. Robson
Cornell University

Summary

When the class boundaries used in constructing a chi-square goodness-of-fit statistic are predetermined and the unknown parameters are estimated by maximum likelihood from the ungrouped data, the resulting statistic does not have a limiting X^2-distribution but instead is asymptotically distributed as a linear function of chi-square variables. The same result applies in the more realistic and useful case where only the number of classes and their probability content are predetermined. It is shown here that in both of the above cases the quadratic form of the asymptotic multinormal conditional distribution of the class frequencies given the parameter estimates can be used to test the goodness-of-fit. This statistic does have a limiting X^2-distribution and the degrees of freedom are only one less than the number of classes after grouping, regardless of the number of parameters estimated.

1. Introduction

The classical procedure for testing whether a sample $x_1, \cdots, x_n$ is obtained from a specified univariate parametric family $f(x;\theta)$, such as Poisson or Normal, employs a statistic measuring goodness-of-fit between the observed (ν_i) and expected (np_i) numbers of observations falling in the r predetermined classes. If $f(x;\theta)$ involves unknown parameters $\theta = (\theta_1, \cdots, \theta_s)$ these can be estimated as

Paper No. BU-388-M in the Biometrics Unit Mimeograph Series, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14850.

functions of ν_i using the maximum likelihood method or minimum χ^2-procedure to obtain estimates $\tilde{p}_i = \tilde{p}_i(\nu_1, \cdots, \nu_r)$ of class probabilities $p_i (i = 1, \cdots, r)$. Under certain regularity conditions (cf. Cramér (1946)) the goodness-of-fit statistic

$$\tilde{R} = \sum_{i=1}^{r} \frac{(\nu_i - n\tilde{p}_i)^2}{n\tilde{p}_i} \tag{1.1}$$

is then asymptotically distributed as χ^2 with r-s-1 degrees of freedom (χ^2_{r-s-1}, briefly). However, if the original observations $x_1, \cdots, x_n$ are available and if the class frequencies $\nu_1, \cdots, \nu_{r-1}$ are not a statistically sufficient reduction of $x_1, \cdots, x_n$, then more efficient estimators of p_i are available, such as maximum likelihood estimators $\hat{p}_i = p_i(\hat{\theta})$ obtained by maximizing the likelihood of $x_1, \cdots, x_n$ with respect to θ. Chernoff and Lehmann (1954) have shown that the statistic thus constructed

$$X^2 = \sum_{i=1}^{r} \frac{(\nu_i - n\hat{p}_i)^2}{n\hat{p}_i} \tag{1.2}$$

is asymptotically distributed as a linear function of chi-square variables, $X^2 \xrightarrow{\mathcal{L}} y_1^2 + \cdots + y_{r-s-1}^2 + \lambda_1 y_{r-s}^2 + \cdots + \lambda_s y_{r-1}^2$, where y_i are independent standard normal variables and the λ's, constrained by $0 \leq \lambda_i < 1$, may depend on the s unknown parameters $\theta_1, \cdots, \theta_s$.

Chernoff and Lehmann (1954) considered only the case where the class boundaries are predetermined. Subsequently, A. R. Roy (1956) and Watson (1957, 1958) independently showed that this same result applies in the more realistic and useful case where only the number of classes, r, and the $\hat{p}_i$ are predetermined; the class boundaries are then functions of $\hat{\theta}$. Watson (1958) concludes that if the parameters involved are those of location and scale, the asymptotic distribution of (1.2) is independent of parameters. Moore (1971) shows that this asymptotic distribution is dependent on the functional form of $f(x;\theta)$ and he tabulates the percentile points of the asymptotic distribution of (1.2) when $f(x;\theta)$ is the normal distribution with unknown mean and variance.

The main result of this paper is to show that, in the case of predetermined class boundaries and also in the case where class boundaries are functions of estimates of θ, there exists a goodness-of-fit statistic which is asymptotically distributed as χ^2_{r-1}. This is accomplished by considering the conditional distribution of the class frequencies $\nu_1, \cdots, \nu_{r-1}$ which, conditioned on $\hat{\theta}$, is asymptotically nonsingular multinormal. The quadratic form $Q_{r-1}(\nu;\hat{\theta})$ of this asymptotic conditional distribution is then asymptotically distributed as χ^2_{r-1}. If $f(x;\theta)$ belongs to the exponential family then this quadratic form does not depend on θ; in other cases the substitution of $\hat{\theta}$ for θ still gives $Q_{r-1}(\nu;\hat{\theta},\theta) \xrightarrow{\mathcal{L}} \chi^2_{r-1}$. Structurally, this statistic takes the form

$$Q_{r-1}(\nu;\hat{\theta}) = X^2 + Y^2 \tag{1.3}$$

where X^2 is the same as (1.2) and Y^2 is a function of the estimated information matrices $\tilde{J}$ and $\hat{J}$ (cf. Chernoff and Lehmann (1954)) and is asymptotically distributed as $(1 - \lambda_1)y^2_{r-s} + \cdots + (1 - \lambda_s)y^2_{r-1}$. Corresponding results are obtained when the maximum likelihood estimators are replaced by other asymptotically normal and consistent estimators.

2. Examples

We consider two examples of tests of goodness-of-fit, (i) a binomial with three classes where the first two are grouped and (ii) a normal distribution with unknown mean and known variance with two class intervals determined by the sample mean.

Binomial:

Let $\nu = (\nu_0, \nu_1, \nu_2)$, $\nu_0 + \nu_1 + \nu_2 = n$, denote the frequency distribution in a sample from some probability distribution $p = (p_0, p_1, p_2)$ on the integers 0, 1, and 2; and consider the problem of testing goodness-of-fit to the binomial family

$$p_x(\theta) = \binom{2}{x}\theta^x(1 - \theta)^{2-x}, \quad 0 \le \theta \le 1.$$

The conventional approach gives

$$X^2 = \sum_{x=0}^{2} \frac{\left(\nu_x - np_x(\hat{\theta})\right)^2}{np_x(\hat{\theta})}, \quad \hat{\theta} = \frac{\nu_1 + 2\nu_2}{2n} \tag{2.1}$$

which is asymptotically X_1^2. Suppose, however, that one of the observed class frequencies is extremely small, say $\nu_0 = 0$; convention would then dictate that the 0 and 1 classes be combined to give

$$X_g^2 = \frac{\left(\nu_0 + \nu_1 - np_0(\hat{\theta}) - np_1(\hat{\theta})\right)^2}{np_0(\hat{\theta}) + np_1(\hat{\theta})} + \frac{\left(\nu_2 - np_2(\hat{\theta})\right)^2}{np_2(\hat{\theta})} = \frac{(\nu_2 - n\hat{\theta}^2)^2}{n\hat{\theta}^2(1 - \hat{\theta}^2)} . \tag{2.2}$$

The loss of one degree of freedom due to the grouping now creates an impasse which is clarified but not overcome by the results of Chernoff and Lehmann (1954); X_g^2 nominally has zero degrees of freedom but converges in law to some fraction $\lambda(\theta)$ of a chi-square variable on one degree of freedom.

This particular impasse was resolved by H. Levene (1949) who noted that for a fixed value of the sufficient statistic $\hat{\theta}$ the conditional distribution of any of the three class frequencies, say ν_2, is asymptotically normal. This distribution

$$P(\nu_2 | 2n\hat{\theta} = s) = \frac{n!}{(\nu_2 - s + n)!(s - 2\nu_2)!s!} \; \frac{2^{s-2\nu_2}}{\binom{2n}{s}} ,$$

where $s = \nu_1 + 2\nu_2$, is thus approximated by a normal distribution with moments

$$E(\nu_2|\hat{\theta}) = n\hat{\theta}^2 - \frac{n\hat{\theta}(1 - \hat{\theta})}{2n - 1} = n\hat{\theta}^2 + o_p(1)$$

$$\mathrm{Var}(\nu_2|\hat{\theta}) = n\hat{\theta}^2(1 - \hat{\theta})^2 + o_p(1) .$$

The test statistic

$$Q_1(\nu;\hat{\theta}) = \frac{\left(\nu_2 - E(\nu_2|\hat{\theta})\right)^2}{\mathrm{Var}(\nu_2|\hat{\theta})} = \frac{(\nu_2 - n\hat{\theta})^2}{n\hat{\theta}^2(1 - \hat{\theta})^2} + o_p(1) \qquad (2.3)$$

is then approximately χ_1^2. Exact critical values of this particular test statistic have been tabulated by C. Vithayasai (1971) and compared with nominal critical values, showing that chi-square approximation is usable even in small samples with extreme values of $\hat{\theta}$. Note that this approach of Levene (1949) circumvents the distasteful matter of grouping to eliminate small class frequencies.

Asymptotic normality of the conditional distribution of ν_2 loosely follows from the multinomiality of ν; since the joint distribution of ν_1 and ν_2 is asymptotically bivariate normal then so also is the joint density of ν_2 and S. The quadratic form of the resulting conditional normal distribution of ν_2,

$$Q_1(\nu;\hat{\theta},\theta) = \left[\nu_2 - E(\nu_2) - \frac{\mathrm{Cov}(\nu_2,S)}{\mathrm{Var}(S)}\left(s - E(S)\right)\right]^2 \Bigg/ \left[\mathrm{Var}(\nu_2) - \frac{\mathrm{Cov}(\nu_2,S)}{\mathrm{Var}(S)}\right]^2$$

$$= \frac{\left[\nu_2 - n\theta^2 - \theta(s - 2n\theta)\right]^2}{n\theta^2(1 - \theta)^2}$$

where s is the value assumed by S in the sample, depends on θ. But

$$Q_1(\nu;\hat{\theta}) = Q_1(\nu;\hat{\theta},\hat{\theta}) + o_p(1)$$

and hence all three statistics $Q_1(\nu;\hat{\theta},\theta)$, $Q_1(\nu_1;\hat{\theta})$, and $Q_1(\nu;\hat{\theta},\hat{\theta})$ are asymptotically identically distributed as χ_1^2. This resolution of the impasse is summarized in the relation

$$X_g^2 = \frac{1 - \hat{\theta}}{1 + \hat{\theta}} Q_1(\nu;\hat{\theta},\hat{\theta}) = \lambda(\hat{\theta})Q_1(\nu;\hat{\theta},\hat{\theta})$$

or

$$Q_1(\nu;\hat{\theta},\hat{\theta}) = X_g^2 + \big(1 - \lambda(\hat{\theta})\big)Q_1(\nu;\hat{\theta},\hat{\theta}).$$

We have thus added to $X_g^2 \overset{\mathcal{L}}{\to} \lambda\, X_1^2$, a statistic $Y^2 = \big(1 - \lambda(\hat{\theta})\big)Q_1(\nu;\hat{\theta},\hat{\theta}) \overset{\mathcal{L}}{\to} (1 - \lambda)X_1^2$.

Normal distribution:

H_o: $x_1, \cdots, x_n$ are n independent observations from a Normal distribution with unknown mean μ and known variance 1.

Let $\bar{x}$ denote the sample mean, let the class intervals be $(-\infty,\bar{x})$ and $(\bar{x},\infty)$, and let ν_1 be the number of observations less than $\bar{x}$. The conditional density of x_1 given $\bar{x}$ is

$$f_n(x_1|\bar{x}) = \frac{1}{\sqrt{2\pi(1 - 1/n)}} e^{-(x_1-\bar{x})^2/2(1-1/n)} . \tag{2.4}$$

Let $P_1 = P(x_1 < \bar{x}|\bar{x})$ and $P_{11} = P(x_1 < \bar{x},\ x_2 < \bar{x}|\bar{x})$. For a fixed $\bar{x}$ the conditional distribution of ν_1 can be approximated for large n by normal distribution with $E(\nu_1|\bar{x}) = nP_1$ and $\mathrm{Var}(\nu_1|\bar{x}) = n(P_1 - P_{11}) + n^2(P_{11} - P_1^2)$. The quadratic form

$$Q_1(\nu;\bar{x}) = \big(\nu_1 - E(\nu_1|\bar{x})\big)^2/\mathrm{Var}(\nu_1|\bar{x}) \tag{2.5}$$

is asymptotically distributed as X_1^2. To compute (2.5) we need only to evaluate the conditional mean and variance of ν_1 correct up to terms $O_p(\frac{1}{n})$. For this purpose we can approximate (2.4) using Taylor's expansion for factors involving terms $(1 - 1/n)$ and ignoring terms $o_p(\frac{1}{n})$. The first approximation of (2.4) yields

$$f_n(x_1|\bar{x}) = \frac{1}{\sqrt{2\pi}} (1 + \frac{1}{2n}) \exp\Big(- \frac{(x_1 - \bar{x})^2}{2} (1 + \frac{1}{n})\Big) . \tag{2.6}$$

Let $f = f(x_1;\bar{x}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_1 - \bar{x})^2}{2}\right)$. Then $\frac{d^2f}{d\bar{x}^2} = \left(-1 + (x_1 - \bar{x})^2\right)f$. Approximating (2.6) again, ignoring terms $o_p(\frac{1}{n})$,

$$f_n(x_1|\bar{x}) = f - \frac{1}{2n}\frac{d^2f}{d\bar{x}^2} \ . \tag{2.7}$$

Using (2.7) we obtain, correct up to terms $O(\frac{1}{n})$, $E(\nu|\bar{x}) = n/2$, $\mathrm{Var}(\nu|\bar{x}) = n(\frac{1}{4} - \frac{1}{2\pi})$. Then (2.5) becomes

$$Q_1(\nu;\bar{x}) = (\nu_1 - \tfrac{n}{2})^2/n(\tfrac{1}{4} - \tfrac{1}{2\pi}).$$

Writing $\lambda = 1 - \frac{2}{\pi}$ and $X^2 = (\nu_1 - \frac{n}{2})^2/\frac{n}{4}$, we have

$$Q_1(\nu;\bar{x}) = X^2 + \frac{(1-\lambda)}{\lambda} X^2 \ .$$

3. General Case

In this section, we consider two cases---(i) predetermined class intervals and (ii) variable class intervals. Throughout the discussion the subscripts i and h will assume values 1, ⋯, r, while j and k assume the values 1, ⋯, s.

3.1 Predetermined Class Boundaries

Use of Maximum Likelihood Estimates

Let $X = (x_1, \cdots, x_n)$ be n independent observations with common density $f(x;\eta)$. Let $N(\theta)$ denote a neighborhood of the true but unknown θ. We assume that f satisfies

(a) for almost all x, the derivatives

$$\frac{\partial \log f}{\partial \eta_j}, \frac{\partial^2 \log f}{\partial \eta_j \partial \eta_k}, \frac{\partial^3 \log f}{\partial \eta_j \partial \eta_k \partial \eta_m}$$

exist for every η in the closure of $N(\theta)$.

(b) if $\eta \epsilon N(\theta)$

$$\left| \frac{\partial f}{\partial \eta_j} \right| < F(x), \quad \left| \frac{\partial^2 f}{\partial \eta_j \partial \eta_k} \right| < F(x), \quad \left| \frac{\partial^3 \log f}{\partial \eta_j \partial \eta_k \partial \eta_m} \right| < H(x)$$

where $F(x)$ is finitely integrable and $E\Big(H(x)\Big) < M$, where M is independent of η.

(c) if $\eta \epsilon N(\theta)$, the matrix $\left(E(\frac{\partial \log f}{\partial \eta_j} \frac{\partial \log f}{\partial \eta_k})\right)$ is finite and positive definite.

Let the likelihood function be given by

$$L(X,\eta) = \prod_{\alpha=1}^{n} f(x_\alpha; \eta).$$

From the above conditions it follows that for $\eta \epsilon N(\theta)$

$$\frac{1}{n} \log L(X,\eta) = \frac{1}{n} \log L(X;\theta) + A(\theta - \eta) + \frac{1}{2}(\eta - \theta)'B(\eta - \theta) + |\eta|^3 o_p(1)$$

where A is the vector whose j^{th} component is

$$A_j = \frac{1}{n} \sum_{\alpha=1}^{n} \frac{\partial \log f(x_\alpha; \theta)}{\partial \theta_j}$$

and B is the matrix whose $(j,k)^{th}$ term is

$$B_{jk} = \frac{1}{n} \sum_{\alpha=1}^{n} \frac{\partial^2 \log f(x_\alpha; \theta)}{\partial \theta_j \partial \theta_k} .$$

If the true value of the parameter is given by $\eta = \theta$, then the asymptotic distribution of $\sqrt{n}\, A$ is normal with mean θ and covariance matrix $J = \left(E(\frac{\partial \log f}{\partial \theta_j} \frac{\partial \log f}{\partial \theta_k})\right)$ where J is the positive definite information matrix and $B \to -J$ in probability.

Further we assume that the sample space is divided into r mutually exclusive and exhaustive subsets and suppose that the probability of obtaining a result belonging to the i^{th} group is $p_i = p_i(\theta) = p_i(\theta_1, \cdots, \theta_s)$ $(s < r)$. The $p_i(\eta) = p_i(\eta_1, \cdots, \eta_s)$ satisfy the conditions on pages 426-427 of Cramér (1946)--namely, that

(i) $\displaystyle\sum_{i=1}^{r} p_i(\eta) = 1$

(ii) $p_i(\eta) > c^2 > 0$ for all $i = 1, \cdots, r$

(iii) every p_i has continuous derivatives

$$\frac{\partial p_i}{\partial \eta_j}, \quad \frac{\partial^2 p_i}{\partial \eta_j \partial \eta_k}$$

(iv) the matrix $\left(\frac{\partial p_i}{\partial \eta_j}\right)$ $i = 1, \cdots, r$; $j = 1, \cdots, s$, is of rank s.

Let ν_i denote the number of x's belonging to the i^{th} group which occur in a sequence of n repetitions of the experiment so that $\sum_{i=1}^{r} \nu_i = n$.

Let $z = \left(g_1(x_\alpha), \cdots, g_r(x_\alpha)\right)$ where $g_i(x_\alpha) = 1$ if the α^{th} observation falls in the i^{th} cell and 0 otherwise. Applying the Central limit theorem to the vector

$$\left(g_1(x_\alpha), \cdots, g_{r-1}(x_\alpha), \frac{\partial \log f}{\partial \theta_1}, \cdots, \frac{\partial \log f}{\partial \theta_s}\right) \tag{3.1.1}$$

we see that the joint density of

$$\sqrt{n}\left(\frac{\nu_1}{n}, \cdots, \frac{\nu_{r-1}}{n}, A_1, \cdots, A_s\right)$$

is asymptotically normal with mean vector

$$\sqrt{n}\,(p_1, \cdots, p_{r-1}, 0, \cdots, 0)$$

and the covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where $\Sigma_{11} = \Big(p_i(\delta_{ij} - p_j)\Big)$, $\delta_{ij} = 1$ or 0 according as $i = j$ or $i \neq j$, $\Sigma_{12} = \Sigma_{21}' =$ $\Big(\mathrm{cov}(\frac{\nu_i}{\sqrt{n}}, \sqrt{n}\, A_j)\Big)$ and $\Sigma_{22} = \Big(\mathrm{cov}(\sqrt{n}\, A_j, \sqrt{n}\, A_k)\Big)$. Here

$$\begin{aligned}
\mathrm{Cov}\Big(\frac{\nu_i}{\sqrt{n}}, \sqrt{n}\, A_j\Big) &= E\Big(\frac{1}{n} \sum_{\alpha=1}^{n} g_i(x_\alpha) \frac{\partial \log f(x_\alpha;\theta)}{\partial \theta_j}\Big) \\
&= E\Big(\frac{\partial \log f(x_\alpha;\theta)}{\partial \theta_j} \Big| \, g_i(x_\alpha) = 1\Big) P(g_i(x_\alpha) = 1) \\
&= \int_{\{x: g_i(x)=1\}} \frac{\partial f(x;\theta)}{\partial \theta_j}\, dx = \frac{\partial p_i}{\partial \theta_j}
\end{aligned}$$

$$\mathrm{Cov}\Big(\sqrt{n}\, A_j, \sqrt{n}\, A_k\Big) = E\Big(\frac{\partial \log f(x;\theta)}{\partial \theta_j} \frac{\partial \log f(x;\theta)}{\partial \theta_k}\Big) .$$

Now the asymptotic conditional distribution of $\Big(\frac{\nu_1}{\sqrt{n}}, \cdots, \frac{\nu_{r-1}}{\sqrt{n}}\Big)$ given $\sqrt{n}(A_1, \cdots, A_s)$ is multivariate normal with mean vector

$$\sqrt{n}\, p + \sqrt{n}\, \Sigma_{12}\Sigma_{22}^{-1} A \tag{3.1.2}$$

and the covariance matrix

$$V = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = (v_{ih}) . \tag{3.1.3}$$

The quadratic form of this asymptotic conditional distribution

$$Q_{r-1}(\nu;\theta) = \frac{1}{n}(\nu - np - n\Sigma_{12}\Sigma_{22}^{-1}A)'V^{-1}(\nu - np - n\Sigma_{12}\Sigma_{22}^{-1}A) \tag{3.1.4}$$

is asymptotically distributed as χ^2_{r-1}.

For each fixed i ($i = 1, \cdots, r-1$) the i^{th} component of (3.1.2) is

$$(\nu_i - np_i - n \sum_{j,k} b^{jk} A_k \frac{\partial p_i}{\partial \theta_j}) \tag{3.1.5}$$

where $(b^{jk}) = \Sigma_{22}^{-1}$. The covariance matrix of $n\Sigma_{12}\Sigma_{22}^{-1}A$ is $n\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = n(c_{ih})$. Since $E\left(\frac{\partial \log f}{\partial \theta_j}\right) = 0$, by Khintchine's theorem $A_j \to 0$ in probability for each j. Further, we have $P\left(|n \sum_{j,k} b^{jk} A_k \frac{\partial p_i}{\partial \theta_j}| \geq \frac{\gamma\sqrt{n}}{2}\right) < \frac{4c_{ii}}{\gamma^2}$ and $P\left(|\nu_i - np_i - n \sum_{j,k} b^{jk} A_k \frac{\partial p_i}{\partial \theta_j}| \geq \frac{\gamma\sqrt{n}}{2}\right) < \frac{4v_{ii}}{\gamma^2}$. The probability that we have $|n \sum_{j,k} b^{jk} A_k \frac{\partial p_i}{\partial \theta_j}| \geq \frac{\gamma\sqrt{n}}{2}$ for at least one value of i is less than $\frac{4\Sigma c_{ii}}{\gamma^2}$ and conversely with a probability greater than $1 - \frac{4\Sigma c_{ii}}{\gamma^2}$ we have

$$|n \sum_{j,k} b^{jk} A_k \frac{\partial p_i}{\partial \theta_j}| \leq \frac{\gamma\sqrt{n}}{2} \quad \text{for all i.} \tag{3.1.6}$$

Similarly with a probability greater than $1 - \frac{4\Sigma v_{ii}}{\gamma^2}$ we have

$$|\nu_i - np_i - n \sum_{j,k} b^{jk} A_k \frac{\partial p_i}{\partial \theta_j}| < \frac{\gamma\sqrt{n}}{2} \quad \text{for all i.} \tag{3.1.7}$$

Therefore $P\left(|\nu_i - np_i| < \gamma\sqrt{n}, \text{ for all i}\right) \geq 1 - \frac{4(\sum_i v_{ii} + \sum_i c_{ii})}{\gamma^2} \to 1$ as $n \to \infty$. We assume that (3.1.6) holds and ν_i satisfy (3.1.7). As in Cramér (1946) we denote γ to be a function of n such that $\gamma \to \infty$ with n, while $\gamma^2/\sqrt{n} \to 0$. We may take $\gamma = n^q$, $0 < q < \frac{1}{4}$. Therefore all the results obtained under these assumptions will be true with a probability equal to $\min\left(1 - \frac{4\sum_i v_{ii}}{\gamma^2}, 1 - \frac{4\sum_i c_{ii}}{\gamma^2}\right)$ which tends to 1 as $n \to \infty$.

<u>Theorem 1</u>: $Q_{r-1}(\nu;\theta)$ of (3.1.4) can be expressed as the sum of two quadratic forms, one of which is positive definite and the other non-negative definite; and

$Q_{r-1}(\nu;\theta)$ is asymptotically distributed as

$$\sum_{i=1}^{r-s-1} y_i^2 + \sum_{i=r-s}^{r-1} \lambda_i y_i^2 + \sum_{i=r-s}^{r-1} (1 - \lambda_i) y_i^2$$

where $0 \le \lambda_i < 1$ and $y_1, \cdots, y_{r-1}$ are independent standard normal variables.

Proof: Let $V^{-1} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} = \Sigma_{11}^{-1} + \left[(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} - \Sigma_{11}^{-1}\right]$. Consider $\Sigma_{11} - (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ which is non-negative definite symmetric matrix because it is the covariance matrix of $\sqrt{n}\ \Sigma_{12}\Sigma_{22}^{-1}A$. Therefore, $(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} - \Sigma_{11}^{-1}$ is non-negative definite matrix. The characteristic roots of $(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} - \Sigma_{11}^{-1}$ are the same as the roots of $|\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \lambda\Sigma_{11}| = 0$ and are all ≥ 0. Since rank $(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) = s$, only s of the λ's are not equal to zero. Further, the roots of $|\Sigma_{11}^{-1}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) - \mu I| = 0$ are the same as the roots of $|(I - \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) - \mu I| = 0$. Since $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is positive definite $\mu_i > 0$ and from the above discussion $\mu_i = 1 - \lambda_i > 0$ which implies $0 \le \lambda_i < 1$.

If we define $Y = P(\nu - np - n\Sigma_{12}\Sigma_{22}^{-1}A)$, where $P'P = V$, then Y is distributed as a multivariate normal with mean vector 0 and covariance matrix I. Therefore $Q_{r-1}(\nu;\theta)$ is distributed as

$$y_1^2 + \cdots + y_{r-s-1}^2 + \sum_{i=r-s}^{r-1} \lambda_i y_i^2 + \sum_{i=r-s}^{r-1} (1 - \lambda_i) y_i^2 = \sum_{i=1}^{r-1} y_i^2$$

where y_i are standard normal variables. Hence, $Q_{r-1}(\nu;\theta) \xrightarrow{\mathcal{L}} \chi^2_{r-1}$.

Since (3.1.2) is a continuous real valued function in $(\theta_1, \cdots, \theta_s)$ and since the maximum likelihood estimate $\hat{\theta}_n = (\hat{\theta}_1, \cdots, \hat{\theta}_s)$ of θ tends to θ in probability (noting that A becomes a zero vector when we substitute $\hat{\theta}$ for θ), we observe that

$\sqrt{n}\, p(\hat{\theta})$ tends in probability to $\sqrt{n}\, p + \sqrt{n}\, \Sigma_{12}\Sigma_{22}^{-1}A$. Further the elements of all the matrices that occur in $Q_{r-1}(\nu;\theta)$ are real continuous functions in θ. Therefore

$$Q_{r-1}(\nu;\hat{\theta}) = \frac{1}{n}(\nu - n\hat{p})'\hat{V}^{-1}(\nu - np)$$

tends, in probability, to $Q_{r-1}(\nu;\theta)$. Since convergence in probability implies convergence in distribution, $Q_{r-1}(\nu;\hat{\theta}) \to \chi^2_{r-1}$.

The computational form of $Q_{r-1}(\nu;\hat{\theta})$ is given by

Theorem 2: $Q_{r-1}(\nu;\hat{\theta}) = X^2 + Y^2$ where $X^2 = \sum_{i=1}^{r} \frac{(\nu_i - n\hat{p}_i)^2}{n\hat{p}_i}$ and

$$Y^2 = \frac{1}{n} \sum_{j,k} \left(\sum_i \frac{(\nu_i - n\hat{p}_i)}{\hat{p}_i} \frac{\partial p_i}{\partial \theta_j}\bigg|_{\theta_j = \hat{\theta}_j} \right) \left(\sum_i \frac{(\nu_i - n\hat{p}_i)}{\hat{p}_i} \frac{\partial p_i}{\partial \theta_k}\bigg|_{\theta_k = \hat{\theta}_k} \right) \hat{a}^{jk} \text{ , where}$$

$\hat{a}^{jk} = -(\hat{\tilde{J}} - \hat{\hat{J}})^{-1}$, $\hat{\tilde{J}}$ and $\hat{\hat{J}}$ are obtained by substituting $\hat{\theta}$ for θ in $\tilde{J} = \left(\sum \frac{1}{p_i} \frac{\partial p_i}{\partial \theta_j} \frac{\partial p_i}{\partial \theta_k}\right)$ and $\hat{J} = \left(E\left(\frac{\partial \log f}{\partial \theta_j} \frac{\partial \log f}{\partial \theta_k}\right)\right)$.

Proof: Since $V^{-1} = \Sigma_{11}^{-1} - \Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \Sigma_{22}\right)^{-1}\Sigma_{21}\Sigma_{11}^{-1}$, noting that $\sum_i p_i = 1$, $\sum_i \frac{\partial p_i}{\partial \theta_j} = 0$ for each j, we have $\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \left(\sum_i \frac{1}{p_i} \frac{\partial p_i}{\partial \theta_j} \frac{\partial p_i}{\partial \theta_k}\right) = \tilde{J}$ and $\Sigma_{22} = \hat{J}$ by definition. $(\nu - n\hat{p})'\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12} = \left(\sum_i \frac{(\nu_i - n\hat{p}_i)}{\hat{p}_i} \frac{\partial p_i}{\partial \theta_1}, \ldots, \sum_i \frac{(\nu_i - n\hat{p}_i)}{\hat{p}_i} \frac{\partial p_i}{\partial \theta_s}\right)\bigg|_{\theta = \hat{\theta}}$.

Remark 1: If the distribution under consideration belongs to Koopman-Darmois family then the dimension of the sufficient statistic is the same as the dimension of θ; and the solution of the maximum likelihood equations will be a function of the sufficient statistic. In this case we can employ the sufficient statistic as conditioning variables.

Remark 2: The foregoing development applies well to tests of goodness-of-fit of continuous distributions and discrete distributions that have support on a countably infinite set of points on the real line. The case of discrete distributions, involving s unknown parameters, that have support on finite number, r, of points on the real line needs to be considered specially. For, consider a goodness-of-fit test of a discrete distribution involving two parameters and four classes with the last two classes being grouped. The development in Chernoff and Lehmann (1954) indicates that the statistic X^2 of (1.2) will be distributed as $\lambda_1 y_1^2 + \lambda_2 y_2^2$, $0 < \lambda_i < 1$, $i = 1, 2$, where y_1 and y_2 are independent standard normal variates. However, the degrees of freedom of the X^2-statistic cannot exceed 1 in this case and this is true for Q also. The following corollary explains what happens in these cases.

Let us assume that the $r - \ell + 1, \cdots, r$ classes are grouped.

Corollary 1: The maximum number of degrees of freedom for the Q statistic is r-s-1 and this is achieved when ℓ is equal to s + 1.

To prove this it is sufficient to examine the roots of $|\tilde{J} - \mu\hat{J}| = 0$ and show that if $\ell < s + 1$ then exactly $\ell - 1$ of the λ's are non-zero and $s - \ell + 1$ of the λ's are zero. If the last ℓ classes are grouped then the (j,k) term of $(\tilde{J} - \hat{J})$ is

$$(\tilde{J} - \hat{J})_{jk} = -\sum_i \frac{1}{p_i} \frac{\partial p_i}{\partial \theta_j} \frac{\partial p_i}{\partial \theta_k} + \frac{1}{\sum_i p_i} \left(\sum_i \frac{\partial p_i}{\partial \theta_j}\right) \left(\sum_i \frac{\partial p_i}{\partial \theta_k}\right)$$

(all the summations extend from $r - \ell + 1$ to r)

$$= - \left[\frac{1}{\sum_i p_i} \sum_i \sum_{\substack{h \\ i<h}} p_i p_h \left(\frac{1}{p_i} \frac{\partial p_i}{\partial \theta_j} - \frac{1}{p_h} \frac{\partial p_h}{\partial \theta_j}\right) \left(\frac{1}{p_i} \frac{\partial p_i}{\partial \theta_k} - \frac{1}{p_h} \frac{\partial p_h}{\partial \theta_k}\right)\right] .$$

Then $(\tilde{J} - \hat{J}) = C'PC$ where

$$C = \begin{bmatrix} \frac{1}{p_{r-\ell+1}} \frac{\partial p_{r-\ell+1}}{\partial \theta_1} - \frac{1}{p_{r-\ell+2}} \frac{\partial p_{r-\ell+2}}{\partial \theta_1} & \cdots & \frac{1}{p_{r-\ell+1}} \frac{\partial p_{r-\ell+1}}{\partial \theta_s} - \frac{1}{p_{r-\ell+2}} \frac{\partial p_{r-\ell+2}}{\partial \theta_s} \\ \frac{1}{p_{r-\ell+1}} \frac{\partial p_{r-\ell+1}}{\partial \theta_1} - \frac{1}{p_{r-\ell+3}} \frac{\partial p_{r-\ell+3}}{\partial \theta_1} & \cdots & \frac{1}{p_{r-\ell+1}} \frac{\partial p_{r-\ell+1}}{\partial \theta_s} - \frac{1}{p_{r-\ell+3}} \frac{\partial p_{r-\ell+3}}{\partial \theta_s} \\ \vdots & \vdots & \vdots \\ \frac{1}{p_{r-1}} \frac{\partial p_{r-1}}{\partial \theta_1} - \frac{1}{p_r} \frac{\partial p_r}{\partial \theta_1} & \cdots & \frac{1}{p_{r-1}} \frac{\partial p_{r-1}}{\partial \theta_s} - \frac{1}{p_r} \frac{\partial p_r}{\partial \theta_s} \end{bmatrix}$$

is a $\frac{\ell(\ell-1)}{2} \times s$ matrix of which only the first $(\ell-1)$ rows are independent and the remaining rows are linear combinations of the first $(\ell-1)$ rows.

$$P = \frac{1}{\sum_i p_i} \operatorname{diag}\left(p_{r-\ell+1}p_{r-\ell+2},\ p_{r-\ell+1}p_{r-\ell+3},\ \cdots,\ p_{r-\ell+1}p_r,\ \cdots,\ p_{r-1}p_r\right)$$

so the rank of $(\tilde{J} - \hat{J})$ is $\ell-1$, if $\ell < s + 1$, and equal to s if $\ell \geq s + 1$. Hence the determinantal equation $|\tilde{J} - \mu\hat{J}| = 0$ has exactly $s - \ell + 1$ of μ's equal to 1. This implies the corresponding λ's in Chernoff and Lehmann (1954) are zero, for $\lambda_i = 1 - \mu_i$.

The minimum value of ℓ for which $(\tilde{J} - \hat{J})$ is an invertible matrix is $s + 1$ (that is, in this case the conditional distribution of $\nu_1, \cdots, \nu_{r-\ell}$ given the maximum likelihood equations is nonsingular multivariate normal) and in this situation Q statistic has maximum number of degrees of freedom, $r - s - 1$. In this case given the particular distribution one can show that $Q_{r-1}(\nu;\hat{\theta}) = \tilde{R}$ of (1.2).

Similarly, if there is an arbitrary grouping of classes and several groupings are made, then looking at the matrix C one can determine the exact number of degrees of freedom for the statistic Q. For example, if m groups of sizes $n_1, \cdots, n_m$ are formed out of r classes and $m \geq s + 1$, then $\sum_{i=1}^{m} n_i - m$ $(> s)$ of the columns of C are independent and $\tilde{J} - \hat{J}$ has full rank s.

Use of Sample Moments

The argument offered in the part---Use of Maximum Likelihood Estimates---can be modified to apply when $\theta_1, \cdots, \theta_s$ are estimated by the method of moments. Let $(x_1, \cdots, x_n)$ be a set of independent observations from a distribution involving s unknown parameters $\theta_1, \cdots, \theta_s$. Further let the first s raw moments of the distribution exist as explicit functions $\alpha_j(\theta_1, \cdots, \theta_s)$ of the unknown parameters. If

$$a_j = \sum_{\beta=1}^{n} \frac{x_\beta^j}{n} \quad j = 1, \cdots, s$$

denote the sample raw moments, then the method of moments consists of equating the values a_j computed from the sample to the hypothetical moments

$$\alpha_j(\theta_1, \cdots, \theta_s) = a_j \quad j = 1, \cdots, s$$

and solving for $\theta_1, \cdots, \theta_s$. Since a_j is the mean of n random variables and if $E(x_\beta^j)$, the j^{th} raw moment, exists then by Khintchine's theorem $a_j \to \alpha_j(\theta_1, \cdots, \theta_s)$ in probability so that a_j is a consistent (and also unbiased) estimator of α_j. Now if the correspondence between $\theta_1, \cdots, \theta_s$ and $\alpha_1, \cdots, \alpha_s$ is one-to-one and inverse functions

$$\theta_j = q_j(\alpha_1, \cdots, \alpha_s) \quad j = 1, \cdots, s \tag{3.1.8}$$

are continuous in $\alpha_1, \cdots, \alpha_s$ then

$$\hat{\theta}_j = q_j(a_1, \cdots, a_s) \quad j = 1, \cdots, s$$

are solutions of (3.1.8) and $q_j(a_1, \cdots, a_s)$ is a consistent estimator of θ_j.

Let $f(x,\theta)$ be the density function (with θ unknown) from which the sample $x_1, \cdots, x_n$ is obtained. We assume $p_i(\theta)$ satisfy the conditions described in section 3.1. Now applying the Central limit theorem to the vector

$$\left(g_1(x_\beta),\ \cdots,\ g_{r-1}(x_\beta),\ x,\ \cdots,\ x^s\right)$$

we see that the joint density of

$$\sqrt{n}\left(\frac{\nu_1}{n},\ \cdots,\ \frac{\nu_{r-1}}{n},\ \frac{\Sigma x_\beta}{n},\ \cdots,\ \frac{\Sigma x_\beta^s}{n}\right)$$

is multivariate normal with mean vector

$$\sqrt{n}\left(p_1,\ \cdots,\ p_{r-1},\ \alpha_1,\ \cdots,\ \alpha_s\right)$$

and covariance matrix

$$\Sigma = \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right]$$

where $\Sigma_{11} = \left(p_i(\delta_{ij} - p_j)\right)$, $\delta_{ij} = 1$ if $i = j$
$= 0$ if $i \neq j$

$$\Sigma_{12} = \left(\mathrm{Cov}\left(\frac{\nu_i}{\sqrt{n}},\ \sqrt{n}\ \Sigma x_\beta^j\right)\right) = \Sigma_{21}',\ \Sigma_{22} = \left(\mathrm{Cov}\left(\sqrt{n}\ \Sigma x_\beta^j,\ \sqrt{n}\ \Sigma x_\beta^k\right)\right)$$

where

$$\mathrm{Cov}\left(\frac{\nu_i}{\sqrt{n}},\ \sqrt{n}\ \Sigma x_\beta^j\right) = \int_{\{x:g_i(x)=1\}} (x_\beta^j - \alpha_j) f(x,\theta)dx = m_{ij}$$

$$\mathrm{Cov}\left(\sqrt{n}\ \Sigma_\beta x_\beta^j,\ \sqrt{n}\ \Sigma_\beta x_\beta^k\right) = \alpha_{j+k} - \alpha_j\alpha_k\ .$$

Now following a similar argument as in section 3.1 and noting that $\sum_{i=1}^{r-1} m_{ij} = -m_{rj}$ for each j, we can write the statistic

$$Q_{r-1}(\nu;\hat{\theta}) = \sum_{i=1}^{r} \frac{(\nu_i - n\hat{p}_i)^2}{n\hat{p}_i} + \frac{1}{n}\sum_j\sum_k \left(\sum_i \frac{\nu_i - n\hat{p}_i}{\hat{p}_i}\, m_{ij}\right)\left(\sum_i \frac{\nu_i - n\hat{p}_i}{\hat{p}_i}\, m_{ik}\right)\hat{a}^{jk}$$

where

$$(\hat{a}^{jk}) = -\left(\left(\sum_{i=1}^{r} \frac{1}{p_i} m_{ij} m_{ik}\right) - (a_{j+k} - a_j a_k)\right)^{-1}$$

and $Q_{r-1}(\nu;\hat{\theta})$ converges to $Q_{r-1}(\nu;\theta)$ in probability and hence in distribution.

3.2 Variable Class Boundaries

Use of Sufficient Statistic

In this section we consider the problem of goodness-of-fit tests of continuous distributions when $f(x;\theta)$ admits a sufficient statistic and the maximum likelihood method of estimation is employed. It is assumed that the class boundaries are uniquely determined as functions of estimates of the parameters in such a manner that the class intervals have predetermined probability content. Let $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_s)$ be the maximum likelihood estimator of θ and $-\infty = z_0(\hat{\theta}) < z_1(\hat{\theta}) < \cdots < z_{r-1}(\hat{\theta}) < z_r(\hat{\theta}) = \infty$ denote the class boundaries such that for predetermined p_i

$$\int_{z_{i-1}(\hat{\theta})}^{z_i(\hat{\theta})} f(x;\hat{\theta})dx = p_i \quad (i = 1, \cdots, r). \tag{3.2.1}$$

Let the interval $\left(z_{i-1}(\hat{\theta}),\ z_i(\hat{\theta})\right)$ be denoted by $\hat{I}_i$. Let $g_i(x_\alpha) = 1$ if $x_\alpha \epsilon \hat{I}_i$, 0 otherwise. Then the number of x's falling in the i^{th} group is given by $\nu_i = \sum_{\alpha=1}^{n} g_i(x_\alpha)$, $i = 1, \cdots, r$. In this representation the asymptotic conditional distribution of $\frac{1}{\sqrt{n}}(\nu_1, \cdots, \nu_{r-1})$ given $\hat{\theta}$, by the central limit theorem applied to the vector $\left(g_1(x_\alpha), \cdots, g_{r-1}(x_\alpha)\right)$, is multinormal with

$$\begin{aligned} E(\nu_i|\hat{\theta}) &= nP_i \\ Var(\nu_i|\hat{\theta}) &= n(P_i - P_{ii}) + n^2(P_{ii} - P_i^2) \\ Cov(\nu_i,\nu_h|\hat{\theta}) &= -nP_{ih} + n^2(P_{ih} - P_iP_h) \end{aligned} \tag{3.2.2}$$

where

$$P_i = P(x_1 \epsilon \hat{I}_i | \hat{\theta}) = \int_{\hat{I}_i} f_n(x_1|\hat{\theta})dx_1$$

$$P_{ih} = P(x_1 \epsilon \hat{I}_i,\ x_2 \epsilon \hat{I}_h | \hat{\theta}) = \int_{\hat{I}_i}\int_{\hat{I}_h} f_n(x_1,x_2|\hat{\theta})dx_1 dx_2 \tag{3.2.3}$$

$f_n(x_1|\hat{\theta})$ is the marginal conditional density function of x_1 given $\hat{\theta}$ and similarly $f_n(x_1,x_2|\hat{\theta})$ is joint conditional density of x_1,x_2 given $\hat{\theta}$. The conditional density function is difficult to calculate and depends on the particular family under consideration; this difficulty can be obviated by noting that we need only evaluate P_i and P_{ih} correct up to $O_p(\frac{1}{n})$ terms.

We assume that $f(x;\theta)$ satisfies the regularity conditions (cf. Cramér, 1946) such that $\hat{\theta}$ is distributed asymptotically multinormal with mean vector and covariance matrix

$$\theta + o(\tfrac{1}{n}) \quad \text{and} \quad \frac{\hat{J}^{-1}}{n} + o(\tfrac{1}{n})$$

respectively. With this assumption we shall give an approximation to $f_n(x_1|\hat{\theta})$ in terms of $f(x;\hat{\theta})$, where $f(x;\hat{\theta})$ is obtained by substituting $\hat{\theta}$ for θ in $f(x;\theta)$. To this end we prove the following three lemmas. First we reproduce, lemma 1, from Feller (1966, pp. 218-219).

For $n = 1, 2, \cdots$ consider a family of distributions $F_{n,\theta}$ with expectation θ and variance $\sigma_n^2(\theta)$ where θ is a parameter varying in a finite or infinite interval.

<u>Lemma 1</u>: If $\sigma_n^2(\theta) \to 0$ then $E_{n,\theta}(u) \to u(\theta)$ for every bounded continuous function u. The convergence is uniform in every subinterval in which $\sigma_n^2(\theta) \to 0$ uniformly and u is uniformly continuous.

Let $g(n)$ be a positive function defined for all positive integers n. Under the same conditions as in lemma 1, we prove

Lemma 2: If $\frac{|E_{n,\theta}(u)|}{g(n)} \to 0$ uniformly in θ then $\frac{|u(\theta)|}{g(n)} \to 0$ as n increases.

Proof: Suppose not. Then there exists a θ and $\epsilon > 0$ such that $|u(\theta)| > \epsilon g(n)$ for infinitely many n. That is,

$$\epsilon < \frac{|u(\theta)|}{g(n)} = \frac{|u(\theta) - E_{n,\theta}(u) + E_{n,\theta}(u)|}{g(n)}$$

$$\leq \frac{|u(\theta) - E_{n,\theta}(u)|}{g(n)} + \frac{|E_{n,\theta}(u)|}{g(n)} \leq \frac{\eta}{g(n)} + \frac{|E_{n,\theta}(u)|}{g(n)}$$

or, $|E_{n,\theta}(u)| > \epsilon g(n) - \eta$ for infinitely many n. For n large, since η is arbitrary, this implies $|E_{n,\theta}(u)| \geq \epsilon g(n)$ for infinitely many n which contradicts the conclusion of lemma 1.

Lemmas 1 and 2 can be generalized in an obvious way to the case where θ is a vector consisting of s components. Because of the regularity conditions assumed earlier on $f(x;\theta)$ we have the density function $h(\hat{\theta};\theta)$ of $\hat{\theta}$ satisfying the conditions of lemma 1. Using lemma 2 we prove for any x (= $x_1, x_2, \cdots$, or x_n)

Lemma 3:

$$f_n(x|\hat{\theta}) = f - \frac{1}{2n} \sum_{j,k} \hat{\hat{J}}^{jk} \frac{\partial^2 f}{\partial\hat{\theta}_j \partial\hat{\theta}_k} + o_p(\tfrac{1}{n}) \tag{3.2.4}$$

where $(\hat{\hat{J}}^{jk}) = \hat{\hat{J}}^{-1}$ is obtained by substituting $\hat{\theta}$ for θ in $\hat{J}^{-1}$ and $f = f(x;\hat{\theta})$.

Proof: For large n, using Taylor's expansion around θ we have

$$f(x;\hat{\theta}) = f(x;\theta) + \Sigma(\hat{\theta}_j - \theta_j) \frac{\partial f(x;\theta)}{\partial\theta_j} + \frac{1}{2} \sum_{j,k} (\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k) \frac{\partial^2 f(x;\theta)}{\partial\theta_j \partial\theta_k}$$

$$+ \cdots \tag{3.2.5}$$

Also, $\hat{\hat{J}}^{jk} \to \hat{J}^{jk}$ and $\dfrac{\partial^2 f}{\partial\hat{\theta}_j\partial\hat{\theta}_k} \to \dfrac{\partial^2 f(x;\theta)}{\partial\theta_j\partial\theta_k}$. Using (3.2.5) we see that

$$\int \Big(f_n(x|\hat{\theta}) - f(x;\hat{\theta}) + \frac{1}{2n} \sum_{j,k} \hat{\hat{J}}^{jk} \frac{\partial^2 f}{\partial\hat{\theta}_j\partial\hat{\theta}_k} + o_p(\tfrac{1}{n})\Big) h(\hat{\theta};\theta)d\hat{\theta} = o_p(\tfrac{1}{n}).$$

And by lemma 2, this shows that $f_n(x|\hat{\theta})$ can be approximated, uniquely, by the right side of (3.2.4).

Using this approximation for $f_n(x|\hat{\theta})$ we can evaluate P_i and P_{ih} ($i,h = 1, \cdots, r-1$) correct up to $o_p(\frac{1}{n})$ terms. As in Watson (1958) we introduce the operator $\partial' = (\frac{\partial}{\partial\theta_1}, \cdots, \frac{\partial}{\partial\theta_s})$, a row vector, with the convention that it only operates on $f_i = f(x_i;\hat{\theta})$ ($i = 1, 2$) but not on $\hat{\hat{J}}^{-1}$. We obtain

$$P_i = p_i - \frac{1}{2n} \int_{\hat{I}_i} \partial'\hat{\hat{J}}^{-1}\partial f_1 dx_1$$

$$P_{ih} = \int_{\hat{I}_i}\int_{\hat{I}_h} f_1 f_2 dx_1 dx_2 - \frac{1}{2n}\int_{\hat{I}_i}\int_{\hat{I}_h} \partial'\hat{\hat{J}}^{-1}\partial f_1 f_2 dx_1 dx_2 \tag{3.2.6}$$

$$= p_i p_h - \frac{1}{2n}\Big[p_i \int_{\hat{I}_h} \partial'\hat{\hat{J}}^{-1}\partial f_2 dx_2 + 2\int_{\hat{I}_i} \partial' f_1 dx_1 \hat{\hat{J}}^{-1} \int_{\hat{I}_h} \partial f_2 dx_2 + p_h \int_{\hat{I}_i} \partial'\hat{\hat{J}}^{-1}\partial f_1 dx_1\Big].$$

Thus, asymptotically,

$$E\Big(\frac{1}{\sqrt{n}}\,\nu_i|\hat{\theta}\Big) = \sqrt{n}\;p_i$$

$$\mathrm{Var}\Big(\frac{1}{\sqrt{n}}\,\nu_i|\hat{\theta}\Big) = p_i(1 - p_i) - \int_{\hat{I}_i} \partial' f_1 dx_1 \hat{\hat{J}}^{-1} \int_{\hat{I}_i} \partial f_2 dx_2 \tag{3.2.7}$$

$$\mathrm{Cov}\Big(\frac{1}{\sqrt{n}}\,\nu_i, \frac{1}{\sqrt{n}}\,\nu_h|\hat{\theta}\Big) = -p_i p_h - \int_{\hat{I}_i} \partial' f_1 dx_1 \hat{\hat{J}}^{-1} \int_{\hat{I}_h} \partial f_2 dx_2 .$$

Writing $u_{ij} = \int_{\hat{I}_i} \frac{\partial f_1}{\partial \hat{\theta}_j} dx_1$ we obtain the conditional distribution of $\frac{\nu'}{\sqrt{n}} = \frac{1}{\sqrt{n}}$ $(\nu_1, \cdots, \nu_{r-1})$ given $\hat{\theta}$ as multinormal with mean vector $\sqrt{n}\, p' = \sqrt{n}\,(p_1, \cdots, p_{r-1})$ and covariance matrix $V = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ where $\Sigma_{11} = \left(p_i(\delta_{ij} - p_h)\right)$ (with δ_{ij} = 1 if $i = j$, $i,h = 1, \cdots, r-1$, 0 otherwise), $\Sigma_{22} = \hat{\hat{J}}$, $\Sigma_{12} = (u_{ij}) = \Sigma_{21}'$. Thus

$$Q_{r-1}(\nu;\hat{\theta}) = \frac{1}{n}(\nu - np)'V^{-1}(\nu - np) \xrightarrow{\mathcal{L}} \chi^2_{r-1} .$$

Since the p_i's are predetermined constants and $\sum_i p_i = 1$, we have

$$0 = \frac{\partial p_i}{\partial \hat{\theta}_j} = \int_{\hat{I}_i} \frac{\partial f}{\partial \hat{\theta}_j} dx + f(z_i(\hat{\theta});\hat{\theta}) \frac{\partial z_i(\hat{\theta})}{\partial \hat{\theta}_j} - f(z_{i-1}(\hat{\theta});\hat{\theta}) \frac{\partial z_{i-1}(\hat{\theta})}{\partial \hat{\theta}_j}$$

or

$$u_{ij} = f(z_{i-1}(\hat{\theta});\hat{\theta}) \frac{\partial z_{i-1}(\hat{\theta})}{\partial \hat{\theta}_j} - f(z_i(\hat{\theta});\hat{\theta}) \frac{\partial z_i(\hat{\theta})}{\partial \hat{\theta}_j}$$

which implies $\sum_{i=1}^{r} u_{ij} = 0$ for $j = 1, \cdots, s$.

Mann and Wald (1942) proved, in the case where there is no estimation of parameters, that the classical χ^2 goodness-of-fit yields an unbiased test if the class probabilities are equal under the null hypothesis. Following this recommendation in our case also, we can select $p_i = \frac{1}{r}$ $(i = 1, \cdots, r)$ and then the test statistic reduces to

$$Q_{r-1}(\nu;\hat{\theta}) = \frac{r}{n}\sum_i (\nu_i - r/n)^2 + \frac{r^2}{n}\left(\sum_{j,k}\left(\sum_i (\nu_i - r/n)u_{ij}\right)\left(\sum_i (\nu_i - r/n)u_{ik}\right)\hat{a}^{jk}\right) \tag{3.2.8}$$

where $(\hat{a}^{jk}) = (\hat{\hat{J}} - \hat{\tilde{J}})^{-1}$, $\hat{\tilde{J}} = (r \sum_i u_{ij}u_{ik})$.

Use of Sample Moments

Lemma 3 holds even if the maximum likelihood estimators $\hat{\theta}$ are replaced by any consistent and asymptotically normal estimators with variance of the form $\frac{c}{n}$. This can be seen by considering the estimators obtained by the method of moments as in 3.1. The distribution of estimates obtained by the method of moments is asymptotically multinormal. The mean of an estimator differs from the true value by a quantity $O(\frac{1}{n})$ and variance is of the form $\frac{c}{n}$. By a simple correction we may often remove the bias and obtain an unbiased estimator $\hat{\theta}$. Since the method of moments gives estimators which are asymptotically sufficient (Le Cam, 1956) satisfying conditions of lemma 1, then using lemma 2 we can prove for any x $(= x_1, x_2, \cdots, \text{or } x_n)$

Lemma 4:

$$f_n(x|\hat{\theta}) = f(x;\hat{\theta}) - \frac{1}{2n} \sum_{j,k} \hat{V}_{jk} \frac{\partial^2 f}{\partial\hat{\theta}_j \partial\hat{\theta}_k} + o_p(\frac{1}{n}) \qquad (3.2.9)$$

where $\frac{1}{n}(\hat{V}_{jk})$ is the estimated covariance matrix of $\hat{\theta}$. Again, $Q_{r-1}(\nu;\hat{\theta})$ can be obtained as before and $Q_{r-1}(\nu;\hat{\theta}) \to Q_{r-1}(\nu;\theta)$ in probability and hence in distribution to χ^2_{r-1}.

4. Some Numerical Examples

4.1 Predetermined Class Boundaries

Binomial Distribution

Consider a multinomial situation with $r + 1$ classes where the i^{th} class probability, $p_i(\theta) = \binom{r}{i}\theta^i(1 - \theta)^{r-i}$, $i = 0, \cdots, r$. If f_i denotes the number of observations falling into the i^{th} class out of a sample of n observations, then

$$P(f_0 = \nu_0, \cdots, f_r = \nu_r) = \frac{n!}{\prod_i f_i!} \theta^{\sum_i if_i} (1 - \theta)^{rn - \sum_i if_i} \prod_i \binom{r}{i}^{f_i}$$

and $\sum_i if_i = S$ is a sufficient statistic for θ. The conditional probability of obtaining $\nu_0, \cdots, \nu_r$ given $S = s$ is

$$P(f_0 = \nu_0, \cdots, f_r = \nu_r | S = s) = \frac{n!}{\prod_i f_i!} \prod_i \binom{r}{i}^{f_i} \Big/ \binom{rn}{s}.$$

For $r = 3$ the case is considered where $i = 0$ and 1 classes are pooled. In this case we get

$$P_1 = (1 - \theta)^2(1 + 2\theta), \quad P_2 = 3\theta^2(1 - \theta), \quad P_3 = \theta^3;$$

$$\frac{dp_1}{d\theta} = -6\theta(1 - \theta), \quad \frac{dp_2}{d\theta} = 3\theta(1 - 3\theta), \quad \frac{dp_3}{d\theta} = 3\theta^2;$$

$$\tilde{J} = 3(2 - 5\theta)/(1 - \theta)(1 + 2\theta), \quad \hat{J} = 3/\theta(1 - \theta) .$$

Then the test statistic from theorem 2 is given by

$$Q_2(\nu;\hat{\theta}) = \sum_i \frac{(\nu_i - n\hat{p}_i)^2}{n\hat{p}_i} + \frac{1}{n}\left[\sum_i \frac{(\nu_i - np_i)^2}{\hat{p}_i} \frac{dp_i}{d\theta} \Bigg|_{\theta = \hat{\theta}}\right] \frac{\hat{\theta}(1 - \hat{\theta})(1 + 2\hat{\theta})}{3(1 + 5\hat{\theta}^2)}$$

which is distributed asymptotically as χ_2^2. The critical region for such an approximate test of nominal size α is

$$C = \left\{(\nu_1, \nu_2, \nu_3) : Q_2(\nu;\hat{\theta}) > \chi^2_{2;\alpha}\right\} \text{ where } P(\chi^2 > \chi^2_{2;\alpha}) = \alpha.$$

For $\hat{\theta} = .6$ the exact size of this conditional test is evaluated and compared to the nominal value for $\alpha = .05$ and $\alpha = .025$.

Table 1: Simulated Sampling Distribution of Q
for samples from Binomial Distribution

Sample size n	$P(\chi_2^2 > \chi_{2;\alpha}^2) = \alpha$ .05	.025
10	.04801	.00667
20	.03588	.02511
30	.04224	.02338
40	.04681	.02148
50	.05140	.02151
60	.04467	.02507

Normal Distribution

Consider the case of a goodness-of-fit test of a normal distribution with unknown mean μ and unknown variance σ^2 and the predetermined classes $(-\infty, -1)$, $(-1, 0)$, $(0, 1)$ and $(1, \infty)$. If $p_i = p_i(\mu, \sigma)$ $(i = 1, 2, 3, 4)$ denote the probability that an observation falls into the i^{th} cell, then

$$p_1 = \int_{-\infty}^{-1} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

with similar expressions for $i = 2, 3, 4$.

$$\frac{\partial p_1}{\partial \mu} = -\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(1+\mu)^2}{2\sigma^2}}, \quad \frac{\partial p_2}{\partial \mu} = \frac{1}{\sqrt{2\pi}\,\sigma}\left(e^{-\frac{(1+\mu)^2}{2\sigma^2}} - e^{-\mu^2/2\sigma^2}\right)$$

$$\frac{\partial p_3}{\partial \mu} = \frac{1}{\sqrt{2\pi}\,\sigma}\left(e^{-\frac{\mu^2}{2\sigma^2}} - e^{-\frac{(1-\mu)^2}{2\sigma^2}}\right), \quad \frac{\partial p_4}{\partial \mu} = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(1-\mu)^2}{2\sigma^2}}$$

$$\frac{\partial p_1}{\partial \sigma^2} = \frac{(1+\mu)}{2\sqrt{2\pi}\ \sigma^3}\, e^{-\frac{(1+\mu)^2}{2\sigma^2}}, \quad \frac{\partial p_2}{\partial \sigma^2} = \frac{1}{2\sqrt{2\pi}\ \sigma^3}\left[\mu\, e^{-\mu^2/2\sigma^2} - (1+\mu)e^{-(1+\mu)^2/2\sigma^2}\right]$$

$$\frac{\partial p_3}{\partial \sigma^2} = \frac{1}{2\sqrt{2\pi}\ \sigma^3}\left[-(1-\mu)e^{-\frac{(1-\mu)^2}{2\sigma^2}} - \mu\, e^{-\mu^2/2\sigma^2}\right], \quad \frac{\partial p_4}{\partial \sigma^2} = \frac{\mu}{2\sqrt{2\pi}\ \sigma^3}\, e^{-\mu^2/2\sigma^2}$$

$$\tilde{J} = \begin{bmatrix} \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \mu}\right)^2 & \sum_i \frac{1}{p_i}\frac{\partial p_i}{\partial \mu}\frac{\partial p_i}{\partial \sigma^2} \\ \sum_i \frac{1}{p_i}\frac{\partial p_i}{\partial \mu}\frac{\partial p_i}{\partial \sigma^2} & \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \sigma^2}\right)^2 \end{bmatrix} \qquad \hat{J} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}$$

Let $y_1, \cdots, y_n$ be a random sample of $N(0,1)$ variates and $z_i = (y_i - \bar{y})/s_1$ where $\bar{y} = \sum_i y_i/n$, $s_1^2 = \sum_i (y_i - \bar{y})^2/n - 1$. Given $\bar{x}$ and s_1 the variates $x_i = \bar{x} + z_i s$ $(i = 1, \cdots, n)$ will have sample mean $\bar{x}$ and sample variance s^2. For $N = 2000$ samples of size 61 each, the statistic $Q_3(\nu;\hat{\theta})$ is computed and compared against χ^2_α (α = .75, .50, .25, .10, .05, .025, .01) where $P[\chi^2_3 > \chi^2_\alpha] = \alpha$. The following table gives the estimates $\hat{\alpha}$ of α conditioned on several different selected values of $\bar{x}$ and s.

Table 2: Simulated Sampling Distribution of Q
for Samples from Normal Distribution with Unknown Mean and Variance

	$P(\chi^2_3 > \chi^2_{3;\alpha}) = \alpha$						
	.75	.50	.25	.10	.05	.025	.01
$\bar{x} = -0.5$, $s^2 = 1.9887$	.7502	.5105	.2375	.094	.043	.025	.0105
$\bar{x} = 0.0$, $s^2 = 1.75999$	.746	.4975	.25	.095	.0465	.019	.0115
$\bar{x} = 2.5$, $s^2 = 2.2276$	.742	.471	.248	.085	.047	.0215	.005

4.2 Variable Class Boundaries

In the following two examples the r classes are selected such that the class boundaries are functions of $\hat{\theta}$ and each class has the same probability content under $f(x;\hat{\theta})$.

Exponential Distribution

$$H_0\colon\ f(x;\theta) = \frac{1}{\theta}e^{-x/\theta} \quad . \ x \geq 0$$
$$= 0 \qquad x < 0\ , \quad \text{unknown.}$$

The sample mean $\bar{x}$ is a sufficient statistic for θ. Let the r class intervals be $(\bar{x}z_{i-1},\bar{x}z_i)$, $i = 1, \cdots, r-1$, with $z_0 = 0$ and $z_r = \infty$. Let z_i's be determined from

$$\int_{\bar{x}z_{i-1}}^{\bar{x}z_i} f(x;\bar{x})dx = \frac{1}{r}$$

which implies $z_i = -\log(1 - \frac{i}{r})$. Let $v_i = \int_{\bar{x}z_{i-1}}^{\bar{x}z_i} \frac{\partial f(x;\bar{x})}{\partial \bar{x}}\,dx = \frac{1}{\bar{x}}\left(z_{i-1}e^{-z_{i-1}} - z_i e^{-z_i}\right)$

and $u_i = \bar{x}v_i$, $(i = 1, \cdots, r)$. After some simplification, we obtain the test statistic as

$$Q_{r-1}(\nu;\bar{x}) = \frac{r}{n}\sum_i(\nu_i - \frac{n}{r})^2 + \frac{r^2}{n}\,\frac{[\sum_i(\nu_i - n/r)u_i]^2}{(1 - r\sum_i u_i^2)}\ .$$

Thus the value of $\bar{x}$ only shifts the class boundaries without affecting the values of the ν's, and hence of $Q_{r-1}(\nu;\bar{x})$. So, without loss of generality, we can take the value of the conditioning variable $\bar{x}$ to be 1. If $y_1, \cdots, y_n$ is a sample of size n from a standard exponential distribution, $x_i = y_i/\bar{y}$ $\ i = 1, \cdots, n$, where $\bar{y} = \sum_i \frac{y_i}{n}$, will have mean $\bar{x} = 1$. Then ν_i = number of x's falling in $(z_{i-1};z_i)$.

3500 samples of size n (= 50, 100) were generated and for each such sample $Q_{r-1}(\nu;\hat{\theta})$, (for r = 4, 6, 8, 10, 12) and X^2 were computed. In tables 3 (n = 50) and 4 (n = 100), the first line for each value of r gives the value $\hat{\alpha}$, the proportion of samples in which $Q_{r-1}(\nu;\hat{\theta})$ exceeds $X^2_{r-1;\alpha}$, and the second line gives the value $\hat{\alpha}_c$, the proportion of samples in which X^2 exceeds $X^2_{r-2;\alpha}$. In table 5 the case r = 2;3 is considered and the simulated sampling distributions of $Q_{r-1}(\nu;\hat{\theta})$, R - the chi-square statistic for goodness-of-fit when the parameter is assumed known (= 1) and that of X^2 are compared. When r = 2, the asymptotic distribution of X^2 is degenerate.

Normal Distribution

$$H_0:\ f(x;\theta) = f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$
$$-\infty < \mu < \infty,\ \sigma > 0 \text{ unknown.}$$

Let $\bar{x}$ and s^2 be the sample mean and variance. Form the r class intervals $(\bar{x} + z_{i-1}s,\ \bar{x} + z_i s)$ i = 1, ⋯, r where $z_0 = -\infty$, $z_r = \infty$, z_i's are determined from the relation

$$\int_{\bar{x}+z_{i-1}s}^{\bar{x}+z_i s} f(x;\bar{x},s^2)dx = \frac{1}{r} \quad i = 1, \cdots, r.$$

Let $$v_{i1} = \int_{\bar{x}+z_{i-1}s}^{\bar{x}+z_i s} \frac{\partial f(x;\bar{x},s^2)}{\partial\bar{x}}\, dx = \frac{1}{s\sqrt{2\pi}}\left(e^{-z_{i-1}^2/2} - e^{-z_i^2/2}\right)$$ and

$$v_{i2} = \int_{\bar{x}+z_{i-1}s}^{\bar{x}+z_i s} \frac{\partial f(x;\bar{x},s^2)}{\partial s^2} = \frac{1}{2s^2\sqrt{2\pi}}\left(z_{i-1}e^{-z_{i-1}^2/2} - e^{-z_i^2/2}\right).$$

Then $\sum_{i=1}^{r} v_{i1} = 0 = \sum_{i=1}^{r} v_{i2}$. Let $u_{i1} = sv_{i1}$, $u_{i2} = s^2 v_{i2}$,

$$\hat{\tilde{J}} = \begin{bmatrix} r\sum_i u_{i1}^2 & r\sum_i u_{i1}u_{i2} \\ r\sum_i u_{i1}u_{i2} & r\sum_i u_{i2}^2 \end{bmatrix}, \qquad \hat{\hat{J}} = \begin{bmatrix} \frac{1}{s^2} & 0 \\ 0 & \frac{1}{2s^4} \end{bmatrix}.$$

After simplifying the expressions one obtains $\hat{a}^{jk}$, $j,k = 1, 2$ of (3.2.8) as

$$\hat{a}^{11} = -\Big(\sum_i ru_{i2}^2 - 2\Big)/D, \quad \hat{a}^{12} = \hat{a}^{21} = r\sum_i u_{i1}u_{i2}/D,$$

$$\hat{a}^{22} = -\Big(r\sum_i u_{i1}^2 - 1\Big)/D, \text{ where } D = \Big(r\sum_i u_{i1}^2 - 1\Big)\Big(r\sum_i u_{i2}^2 - 2\Big) - r^2\Big(\sum_i u_{i1}u_{i2}\Big)^2 .$$

Then

$$Q_{r-1}(\nu;\bar{x},s^2) = \frac{r}{n}\sum_i(\nu_i - n/r)^2 + \frac{r^2}{n}\Big[\Big(\sum_i(\nu_i - n/r)u_{i1}\Big)^2\hat{a}^{11}$$

$$+ 2\Big(\sum_i(\nu_i - n/r)u_{i1}\Big)\Big(\sum_i(\nu_i - n/r)u_{i2}\Big)\hat{a}^{12}$$

$$+ \Big(\sum_i(\nu_i - n/r)u_{i2}\Big)^2\hat{a}^{22}\Big] .$$

Now it is seen that $Q_{r-1}(\nu;\bar{x},s^2)$ is independent of $\bar{x},s^2$ and thus, for the purpose of simulation, one can take the values of $\bar{x}$ and s^2 to be 0 and 1, respectively. As in the case of predetermined class boundaries, the x_i's are generated by taking $\bar{x} = 0$, $s^2 = 1$. Using these values of x's the statistics $Q_{r-1}(\nu;\bar{x},s^2)$ and X^2 are computed. Using y's the statistic $R = \frac{r}{n}\sum_i(m_i - n/r)^2$ where m_i is the number of y's falling in (z_{i-1}, z_i), $i = 1, \cdots, r$ is computed. R becomes the test statistic to test H_o: $f(y;\theta) = (1/\sqrt{2\pi})\exp(-y^2/2)$. Table 6 gives the comparison of simulated sampling distributions of the goodness-of-fit statistics Q, R, X^2. The first line gives $\hat{\alpha}$, the proportion of samples in which $Q_{r-1}(\nu;\bar{x},s^2)$ exceeds $X^2_{r-1;\alpha}$; the second line gives the $\hat{\alpha}_{c\ell}$, the proportion of samples in which R exceeds $X^2_{r-1;\alpha}$; and the third line gives the proportion of samples in which X^2 exceeds $X^2_{r-3;\alpha}$ for $r = 4$, 6, 8, 10, 12.

Table 3. Simulated sampling distribution of goodness-of-fit statistics for samples (3500) of size 50 from an exponential distribution with unknown mean.

r \ α	.975	.95	.90	.80	.70	.50	.30	.20	.10	.05	.025	.01
4	.978	.94943	.90571	.798	.69971	.51543	.30686	.19686	.10257	.05514	.02714	.00771
	1.000	.96943	.96943	.85229	.85229	.55771	.33971	.26543	.12486	.06857	.032	.01286
6	.97257	.94371	.89829	.79743	.69571	.49514	.29971	.184	.08857	.04029	.02057	.00771
	.98343	.95829	.92743	.81714	.73371	.51886	.32486	.20857	.096	.04486	.02114	.00886
8	.982	.95514	.90429	.79800	.69657	.49743	.288	.18714	.09371	.04714	.02429	.00886
	.97886	.96343	.90486	.83571	.70086	.52571	.31571	.19629	.10743	.04886	.02829	.012
10	.97943	.95343	.90771	.80971	.70686	.49	.278	.18714	.09029	.04629	.02571	.01057
	.98143	.96686	.92429	.82143	.73857	.50143	.30371	.19486	.09686	.04914	.02914	.00886
12	.97914	.95457	.90743	.80657	.69943	.48943	.27771	.18486	.090857	.04514	.02486	.00829
	.98829	.95857	.90971	.8	.70657	.51114	.29829	.18486	.09657	.04943	.02857	.01114

α = nominal size: In each cell first line gives the sampling distribution of Q and the second line that of X^2.

Table 4. Simulated sampling distribution of goodness-of-fit statistics Q and X^2 for samples of size 100 from an exponential distribution with unknown mean.

r \ α		.975	.95	.90	.80	.70	.50	.30	.20	.10	.05	.025	.01
4	Q	.98257	.948	.90029	.79429	.70229	.50143	.30286	.20686	.10286	.056	.02571	.00971
	X^2	.99829	.97886	.968	.88229	.80543	.574	.36486	.24171	.12	.05771	.02914	.01029
6	Q	.97371	.944	.90086	.79657	.692	.50143	.29143	.18714	.09	.044	.02229	.008
	X^2	.98486	.96057	.91229	.82314	.73943	.528	.31771	.20486	.09743	.04514	.02429	.00829
8	Q	.97229	.953429	.90086	.80257	.69714	.49114	.30657	.204	.09686	.04771	.02457	.008
	X^2	.98314	.95686	.922	.81857	.72629	.51143	.30771	.21429	.10057	.05229	.02486	.00886
10	Q	.97571	.95	.89743	.80171	.694	.48343	.27657	.185429	.09	.04343	.02029	.00886
	X^2	.984	.95571	.90714	.826	.71971	.50371	.29571	.19057	.09086	.04771	.02514	.00914
12	Q	.97343	.94429	.896	.80457	.70829	.50486	.296	.19886	.09914	.05029	.02286	.00743
	X^2	.97371	.948	.91286	.81343	.72943	.51686	.31543	.20057	.10171	.058	.02057	.00857

Entries 1st line: Proportion of Q's > $X^2_{r-1;\alpha}$; 2nd line: Proportion of X^2 > $X^2_{r-2;\alpha}$ in samples of 3500.

Table 5. Simulated sampling distribution of goodness-of-fit statistics for samples (3500) of size 500 from an exponential distribution with unknown mean.

r	α	.990	.975	.95	.90	.80	.75	.70	.60	.50	.40
3	Q	.99133	.97567	.94967	.89967	.80067	.74267	.701	.595	.49633	.393
	R	.988	.97767	.95233	.89267	.803	.750	.69267	.60067	.50233	.40633
	X^2	1.00	1.00	1.00	.993	.944	.92033	.87667	.79267	.678	.56233
2	Q	.94733	.94733	.94733	.848	.75167	.75167	.66067	.58567	.50067	.42367
	R	.961	.961	.961	.89733	.83167	.76333	.70067	.63433	.50967	.41333

r	α	.30	.25	.20	.10	.05	.025	.02	.01	.005	.001
3	Q	.299	.24333	.192	.09367	.05167	.025	.02133	.01067	.00467	.00067
	R	.29933	.24867	.20633	.10667	.05267	.02833	.02367	.01067	.00433	.0
	X^2	.42767	.35233	.27533	.13333	.06567	.03267	.024	.01267	.00067	.00067
2	Q	.30367	.24333	.19133	.09167	.05433	.02133	.02133	.01167	.00533	.002
	R	.317	.27833	.20633	.10033	.05833	.025	.01833	.01133	.00533	.001

R: Chi-square goodness-of-fit statistic when the parameter is assumed known and equal to 1.

Table 6: Simulated sampling distribution of goodness-of-fit statistics for samples of size 100 from a normal distribution with unknown mean and variance.

r \ α		.975	.95	.90	.80	.70	.50	.30	.20	.10	.05	.025	.01
	Q	.97833	.949	.91533	.80933	.71	.51067	.29267	.18467	.09567	.051	.028	.01067
4	R	.97833	.95567	.89833	.83	.72233	.50967	.30167	.20733	.103	.05	.02867	.012
	X^2	.999	.999	.999	.999	.964	.82267	.52933	.389	.209	.107	.055	.024
	Q	.97733	.95333	.90	.80933	.71533	.50367	.29667	.20433	.09467	.046	.02367	.00833
6	R	.97267	.95967	.913	.811	.70767	.51367	.31267	.197	.10533	.056	.032	.01467
	X^2	.99533	.98433	.957	.90067	.814	.60967	.392	.25167	.13033	.061	.033	.01433
	Q	.97433	.94667	.89767	.79267	.69367	.49067	.28367	.18567	.087	.04167	.026	.01033
8	R	.98133	.95667	.905	.81333	.69633	.50567	.29867	.19967	.10267	.055	.02733	.01133
	X^2	.98833	.97067	.92867	.855	.76567	.55133	.34033	.23367	.11133	.05847	.02767	.01367
	Q	.970	.95067	.899	.78833	.685	.485	.28833	.19133	.09533	.04767	.02367	.00867
10	R	.978	.95733	.910	.81767	.71633	.508	.30667	.19933	.10567	.052	.02733	.012
	X^2	.98233	.96867	.91467	.817	.73333	.54167	.341	.22167	.11033	.05467	.02867	.0113
	Q	.97367	.95033	.89667	.789	.680	.49467	.289	.19667	.10133	.05367	.025	.00967
12	R	.98033	.95733	.914	.817	.707	.50067	.297	.21067	.11467	.05833	.02133	.011
	X^2	.98267	.961	.907667	.815	.725	.52333	.32733	.222	.11933	.05933	.03167	.01367

5. Discussion

The test statistic $Q_{r-1}(\nu;\hat{\theta})$ is admittedly rather cumbersome to compute when the number of parameters to be estimated is more than one; the classical statistic (1.1) shares this difficulty, however, even if there is only one parameter to be estimated. Further, if the classical procedure is modified to accommodate variable class boundaries the parameters must be estimated twice, by solving maximum likelihood equations formed first using ungrouped data and then using grouped data (Watson (1959, section 4); Moore (1970)) while in using $Q_{r-1}(\nu;\hat{\theta})$ θ is estimated only once. In (1.2) θ is also estimated only once but different tables are needed (Moore (1971)) to test different null hypotheses, while $Q_{r-1}(\nu;\hat{\theta})$ is distribution free and X^2 tables can be used in actual execution of the test. Our simulations support a recommendation by Watson (1958) that the number of classes be increased to 10 or more so that the difference between the asymptotic distribution of (1.2) and that X^2_{r-s-1} is not appreciable. When the number of classes cannot be increased (cf. the contribution to the discussion section of Watson (1958) by Chernoff and Lehmann), then $Q_{r-1}(\nu;\hat{\theta})$ becomes an eminently usable alternative to the recommendation of Chernoff and Lehmann (1954) as implemented by Moore (1971). Again the tables in Moore (1971) are quite restricted in terms of degrees of freedom. Computer programs for calculating $Q_{r-1}(\nu;\hat{\theta})$ can be easily written, however, and those employed in our simulation studies are available upon request in the reports BU-363-M and BU-364-M of the Biometrics Unit Mimeograph Series, Cornell University, to test the goodness-of-fit of exponential and normal distributions, respectively.

References

CHERNOFF, H. and LEHMANN, E. L. (1954). The use of maximum likelihood estimates in X^2 tests for goodness-of-fit. Ann. Math. Statist. 25, 579-586.

CRAMÉR, H. (1946). Mathematical Methods of Statistics. Princeton University Press.

FELLER, W. (1966). An Introduction to Probability Theory and Its Application, Vol. II. John Wiley and Sons, Inc.

KAMBHAMPATI, C. (1971). A chi-square statistic for goodness-of-fit tests. Unpublished Ph.D. thesis, Cornell University.

LE CAM, L. (1956). Proceedings of the Third Berkeley Symposium on the Mathematical Statistics and Probability, Vol. 1, 129-156.

LEVENE, H. (1949). On a matching problem arising in genetics. Ann. Math. Statist. 20, 91-94.

MANN, H. B. and WALD, A. (1942). On the choice of the number of class intervals in the application of the chi-square test. Ann. Math. Statist. 13, 306-317.

MOORE, D. S. (1970). A chi-square statistic with random cell boundaries. A paper read at the Joint Statistical Meetings, University of Wyoming.

MOORE, D. S. (1971). A chi-square statistic with random cell boundaries. Ann. Math. Statist. 42, 147-156.

RAO, C. R. (1965). Linear Statistical Inference and Its Applications. John Wiley and Sons, Inc.

ROY, A. R. (1956). On chi-squared statistics with variable intervals. Department of Statistics, Stanford University, Technical Report.

VITHAYASAI, C. (1971). Exact critical values of the Hardy-Weinberg test statistic for two alleles. Biometrics and Computer Science Branch, Department of Fisheries and Forestry, Canada, Technical Report.

WATSON, G. S. (1957). The chi-squared goodness-of-fit test for normal distributions. Biometrika 44, 336-348.

WATSON, G. S. (1958). On chi-square goodness-of-fit tests for continuous distributions. J. R. Statist. Soc., Ser. B, 20, 44-61.

WATSON, G. S. (1959). Some recent results in chi-square goodness-of-fit tests. Biometrics 15, 440-468.